Assimilation of Temperature into an Isopycnal Ocean General Circulation

Model Using a Parallel Ensemble Kalman Filter

Christian L. Keppenne
Science Applications International Corporation
4600 Powder Mill Road
Beltsville, Maryland 20705

Michele M. Rienecker
NASA Seasonal−to−Interannual Prediction Project
Code 971, Laboratory for Hydrospheric Processes
Goddard Space Flight Center, Greenbelt, Maryland 20771

Corresponding author address:

Christian L. Keppenne
Mail Code 971, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771
tel.: 1−301−3457483 fax.: 1−301−6145644
e−mail: clk@janus.gsfc.nasa.gov

Abstract

Temperature data from the Tropical Atmosphere and Ocean (TAO) array are assimilated into the Pacific basin configuration of the Poseidon quasi−isopycnal ocean general circulation model using a multivariate ensemble Kalman filter (MvEnKF) implemented on a massively parallel computer architecture. An assimilation algorithm whereby each processing element (PE) solves a localized analysis problem is used. The algorithm relies on a locally supported error−covariance model to avoid the introduction of spurious long−range covariances associated with small ensemble sizes and to facilitate its efficient parallel implementation on a computing platform with distributed memory.

Each time data are assimilated, multivariate background−error covariances estimated from the phase−space distribution of an ensemble of model states are used to calculate the Kalman gain matrix and the analysis increments. The multivariate covariances are used to compute temperature, salinity and current increments. The layer thicknesses are left unchanged by the analysis. Instead, they are dynamically adjusted by the model between successive analyses.

Independent acoustic Doppler current profiler data are used to assess the performance of the temperature data assimilation. The temperature analyses are also compared to analyses obtained with a univariate optimal interpolation algorithm and to a control run without temperature assimilation.

The results demonstrate that the MvEnKF is both practical and effective for assimilating *in situ* and remotely sensed observations into a high resolution ocean model in a quasi−operational framework.

1. Background

The purpose of the NASA Seasonal−to−Interannual Prediction Project (NSIPP) is to further the utilization of *in situ* and satellite observations for prediction of short term climate phenomena. NSIPP undertakes routine forecasts in a research framework with global coupled ocean−atmosphere−land surface models. The initial implementation has used a univariate optimal interpolation (univariate OI: UOI, e.g., Troccoli et al. 2001) analysis scheme to assimilate temperature data into the Poseidon quasi−isopycnal ocean general circulation model (OGCM, Schopf and Loughe 1995; Konchady et al. 1998; Yang et al. 1999). Like several other ocean data assimilation systems currently in use at other institutions (e.g., Ji and Leetma 1997), the UOI system is based on the assumption that the forecast−error covariances are approximately Gaussian and that the covariances between the temperature−field errors and the salinity−field and current−field errors are negligible.

In addition to the neglect of cross−field covariances, a serious limitation of the UOI is the fact that it does not use any statistical information about the expected inhomogeneous distribution of model errors. Another limitation is that it is based on a steady state error−covariance model which gives the same weight to a unit innovation regardless of how accurate the ocean−state estimate has become as a result of previous analyses. The last−mentioned limitation is partially a result of the failure to provide time−dependent estimates of the model errors.

A multivariate OI (MvOI, Borovikov and Rienecker 2001) data assimilation system has been implemented in response to the first two limitations. Like the UOI, the MvOI relies on a steady state error−covariance model. However, it takes into account that the model errors may be inhomogeneously distributed and uses multivariate cross−field covariances to update the model

state variables when either direct measurements of a prognostic variable or indirect measurements of a diagnostic quantity such as sea surface height (SSH) are assimilated. To simultaneously address all three aforementioned limitations, a multivariate ensemble Kalman filter (multivariate EnKF: MvEnKF) has been developed and implemented on a massively parallel CRAY T3E architecture. According to an algorithm originally proposed by Evensen (1994) and since then implemented for atmospheric and ocean models of varying complexity in a number of different studies [see Keppenne and Rienecker (2001a, hereafter KR01a) for a historical perspective of ensemble Kalman filtering as it pertains to climate modeling applications], the MvEnKF integrates an ensemble of model trajectories and uses the statistics of the ensemble to estimate the model errors. Unlike the UOI and MvOI, the MvEnKF accounts for the fact that past measurements may already have been assimilated, resulting in a more accurate state estimate, and adjusts the structure and amplitude of the gain functions accordingly.

The MvEnKF implementation for the Poseidon model is discussed in detail in a NASA Technical Memorandum (Keppenne and Rienecker 2001b, hereafter KR01b). The parallel algorithms are tested in KR01a.

In the present article, the impact on the simulated equatorial temperature field and on the model currents of using the MvEnKF to assimilate temperature data from the Tropical Atmosphere and Ocean (TAO, McPhaden *et al.* 1998) array is examined by comparison to NOAA/Pacific Marine Environment Laboratory (PMEL) analyses and to acoustic Doppler current profiler (ADCP) data available at four TAO moorings along the Equator. A model run without assimilation and one in which the data are assimilated using the UOI serve as references.

The remainder of this paper is organized as follows. Section 2 contains an overview of the model and methods which are described in depth in KR01b. Section 3 discusses the TAO–temperature assimilation experiment while Section 4 contains the conclusions and lists some outstanding issues.

2. Model and methods

2.1. The Poseidon parallel ocean model

The Poseidon model (Schopf and Loughe, 1995) is a finite–difference reduced–gravity ocean model which uses a generalized vertical coordinate designed to represent a turbulent, well–mixed surface layer and nearly isopycnal deeper layers. Poseidon has been documented and validated in hindcast studies of El Niño (Schopf and Loughe 1995) and has since been updated to include prognostic salinity (*e.g.*, Yang *et al.* 1999). More recently, the model has been used in a numerical study of the surface heat balance along the equator (Borovikov *et al.* 2001) and in an examination of ENSO and its mechanisms during the 1990s (Yuan *et al.* 2001).

Explicit detail of the model, its vertical coordinate representation and its discretization are provided in Schopf and Loughe (1995). The prognostic variables are layer thickness, $h(\lambda,\vartheta,\xi,t)$, temperature, $T(\lambda,\vartheta,\xi,t)$, salinity, $S(\lambda,\vartheta,\xi,t)$, and the zonal and meridional current components, $u(\lambda,\vartheta,\xi,t)$ and $v(\lambda,\vartheta,\xi,t)$, where λ is longitude, ϑ latitude, t time and ξ is a generalized vertical coordinate.

A 20–layer Pacific basin version of the parallel model with uniform $1°$ zonal resolution is used here. The meridional resolution varies between $1/3°$ at the Equator and $1°$ in the

extratropics. A solid boundary is imposed at $45\,{}^\circ S$. There, a no−slip condition is used for the currents and a no−flux condition is used for mass, heat and salt. This version of Poseidon has been parallelized as in Konchady *et al.* (1998) using the same message−passing protocol and 2D horizontal domain decomposition used by Schaffer and Suarez (1998) for the NSIPP−1 atmospheric general circulation model.

There are $173 \times 164 \times 20$ grid boxes, of which 28% are situated over land, resulting in a total of 2.0422×10^6 individual prognostic variables. Although standalone copies of the model can be run usefully on as few as 16 CRAY T3E processors [hereafter referred to as processing elements (PEs)], 256 PEs are used in the MvEnKF ensemble integrations. For a detailed description of the model configuration used in this study, the reader is referred to KR01b.

2.2. The multivariate ensemble Kalman filter

The complete description of the MvEnKF implementation can be found in KR01b. For a more succinct and less technical discussion of the methodology, the reader can consult KR01a. This Section merely lists some distinctive features of the implementation. The details of the parallel algorithm and other issues, such as pre−filtering, incremental analysis updating and the form taken by the measurement functional, are discussed in KR01b.

2.2.1. Conversion from isopycnals to levels

To maintain compatibility with the UOI which interpolates model fields vertically to a series of pre−specified depths (levels) prior to each analysis, needed elements of P^f, the background− error covariance matrix, are calculated on levels rather than on layers. Therefore, the *T*, *S*, *u* and

v fields are transformed from isopycnals to levels and the analysis increments are calculated on the levels before being mapped back to the isopycnals. Sixteen levels are used.

The above scheme results in only *T*, *S*, *u* and *v* being updated. The *h* field is left unchanged by the assimilation. The procedure allows the model to dynamically recalculate *h* from the new density distribution and the target interface buoyancies, as it does at every time step (see Schopf and Loughe 1995).

2.2.2. Ensemble size

With the MvEnKF, PE memory imposes constraints on both the domain decomposition and the ensemble size. The 256 PEs used in this study impose a limit of about 40 ensemble members on the current platform (CRAY T3E−600 with 128MB local RAM per PE). Encouraging results have been obtained with comparably sized ensembles by Mitchell and Houtekamer (2000) with a three−level QG model and by Keppenne (2000) with a two−layer shallow water model. It is shown in the following Sections that the MvEnKF can perform as well as or better than the UOI with even as few as 40 ensemble members (also see KR01a).

2.2.3. Compactly supported error−covariance model

The small ensemble size introduces spurious long−range correlations that must be filtered when the elements of $\boldsymbol{P}^f$ are computed. Following Houtekamer and Mitchell (2001) and a suggestion by Gaspari and Cohn (1999), this filtering is achieved through a Hadamard product $\left(\boldsymbol{A} \circ \boldsymbol{B} \stackrel{\text{def}}{=} \{ \boldsymbol{A} \circ \boldsymbol{b} \}_{ij} = \boldsymbol{A}_{ij} \boldsymbol{B}_{ij} \right)$ of $\boldsymbol{P}^f$ with a local compactly supported correlation function. This function is defined by (4.10) of Gaspari and Cohn (1999). The correlation scales used in the

6

temperature–assimilation experiment of Section 3 are the same as in KR01a (Section 3d therein):

$$l_\lambda = 30^\circ, \qquad l_\vartheta = 15^\circ \quad \text{and} \quad l_z = 500m.$$

The imposition of compact support results in each PE solving a different PE–local analysis problem. This, in turn, results in a more efficient parallel analysis algorithm (KR01b).

2.2.4. Multivariate algorithm

Consider $x_k = [T_k, S_k, u_k, v_k]$ as an element of the state vector, with $1 \leq k \leq n_x$. Here, $n_x = n_\lambda n_\vartheta n_z$, where n_λ, n_ϑ and n_z stand for the number of grid boxes in the zonal, meridional and vertical directions, respectively. Then, the MvEnKF analysis equations can summarily be written as [see KR01b for their derivation]

$$\{C\}_{pq} = c_{pq} = C(r_{pq}), \qquad 1 \leq p \leq n_d, 1 \leq q \leq n_d, \qquad (1a),$$
$$[C \circ L L^T + W] b_i = d - l(x_i^f) + e_i, \qquad 1 \leq i \leq m, \qquad (1b),$$
$$\{F\}_{kp} = \{f_k\}_p = f_{kp} = C(r_{kp}), \qquad 1 \leq k \leq n_x, 1 \leq p \leq n_d, \qquad (1c),$$
$$g_{ik} = L^T b_i \circ f_k, \qquad 1 \leq i \leq m, 1 \leq k \leq n_x, \qquad (1d)$$
$$x_{ik}^a = x_{ik}^f + (x_k - \bar{x}_k) \circ g_{ik}. \qquad 1 \leq i \leq m, 1 \leq k \leq n_x, \qquad (1e)$$

In (1), uppercase boldface symbols represent matrices, lowercase boldface symbols represent vectors and lowercase regular symbols denote scalar variables. The vector, d contains n_d observations, x_i is the *i*th ensemble state vector of length n_x and *m* is the ensemble size. Following Burgers *et al.* (1998), e_i is a white perturbation term included to help prevent the

ensemble from collapsing, resulting in a systematic error underestimation. The superscripts ^{a}

and ^{f} refer to the analyzed state and the forecast, respectively, and $\bar{x}$ is the ensemble mean. The

measurement operator, $l(x)$, relates the state vector to the observations.

Matrix W is the observation–error covariance matrix. The representer matrix, $R = L L^T$, maps

the background–error covariance matrix, P^f, to the error subspace of the measurements.

Matrix C is a correlation matrix, r_{pq} is the distance separating the pth and qth observations and

$C(r)$ is the canonical correlation function mentioned under 2.2.3 [4.10 of Gaspari and Cohn

(1999)]. The elements of b_i are the representer–function amplitudes used to update x_i.

Vector f_k contains idealized correlations between the coordinates of the kth grid box and those

of every observation processed on the current PE. Vector g_{ik} contains weights used to update

the prognostic variables in the kth grid box corresponding to the ith ensemble member, x_{ik}.

2.2.5. System noise representation

Let the evolution of the true state of the climate system, $\mathring{x}(t)$, be represented by

$$\frac{\partial \mathring{x}}{\partial t} = F(\mathring{x}, t) + e(t), \qquad (2)$$

where F represents the model hydrodynamics, physics and forcing, and e combines the model

errors and forcing errors, and is commonly known as system noise or process noise.

In practice, the statistics of *e* are generally unknown and the system noise is the object of parameterization. Motivated by the current lack of information about the model−error statistics, the system−noise model used in this study attempts solely to account for errors in the surface wind stress and heat flux forcing.

Because of the focus on SI variability, the forcing errors are modeled on SI time scales, with each ensemble member being forced by adding a monthly mean perturbation of the monthly mean basic state to analyzed winds from the Spectral Sensor Microwave Imager (SSM/I: Atlas *et al.* 1996). The perturbations are obtained by integrating an ensemble of atmospheric models over observed sea surface temperature data (Reynolds and Smith 1994), but starting each atmospheric integration at a different time. Each member of the atmospheric ensemble used to force the ocean ensemble has the same SI phase, while the spread of the atmospheric ensemble is meant to be representative of the forcing errors (KR01b).

3. TAO−temperature data assimilation

3.1. Experimental setup

In this experiment, a 40−member MvEnKF is run for three months starting January 1, 1993. Daily averaged TAO temperature data are "superobed" as explained in KR01b and assimilated every five days. The central forecast, x_c, defined as the ensemble member closest to $\bar{x}$ in terms of Euclidian distance in phase space after normalizing T, S, u and v by their standard deviations, serves to estimate the true ocean state. The root−mean square deviation from $\bar{x}$ of

9

the individual ensemble members, $\sigma_x \overset{\text{def}}{=} \sqrt{diag(P^f)}$, serves to estimate the forecast–error standard deviation.

The initial state of each ensemble member is obtained from a different model integration spanning ten years and without assimilation. Atmospheric perturbations are added to the same basic monthly mean forcing used for each integration. For reference, a control run without assimilation is run for three months forced with the same SSM/I–wind data used with the MvEnKF and without a forcing perturbation. A third run in which the TAO data are assimilated with the UOI mentioned in Section 1 is started from the same initial condition as the no–assimilation control run and uses the same forcing. The states from the control and UOI runs are referred to as x_{co} and x_{oi}, respectively.

After the three–month period with temperature assimilation, three twelve–month hindcasts forced with climatological data are initialized from x_{co}, x_{oi} and x_c on March 31, 1993. While observed wind forcing is used in the period with temperature assimilation, climatological winds are used in the no–assimilation hindcasts to simulate a forecast scenario. In the absence of a dynamic coupler for the Pacific–basin and atmospheric models, the climatologically forced hindcasts are proxies for coupled–model forecasts made with the global versions of the NSIPP ocean and atmospheric models. The goal is to gain insight into how using the MvEnKF to assimilate the TAO–temperature data into the global version of Poseidon will impact the skill of the coupled SI forecasts.

During the three months with temperature assimilation, the equatorial temperature fields from

x_{co}, x_{oi} and x_{c} are compared to PMEL objective analyses. Also, the zonal currents at

four TAO moorings equipped with current profilers are compared to ADCP data. Observed

temperature anomalies from TAO are used to validate the temperature hindcasts obtained with the

twelve–month runs without assimilation.

3.2. Error estimates

An important aspect of the Kalman filter is that it estimates not only the state of a dynamical

system, but also the uncertainty associated with the state estimates (Gelb 1974; Daley 1991; Ghil

and Malanotte–Rizzoli 1991). With the MvEnKF, the elements of P^{f} required in the analysis

are estimated from the ensemble distribution. Four snapshots through the evolution of the

equatorial–temperature–field portion of σ_{x}, defined above as the square root of the main

diagonal of P^{f}, are shown in Figure 1. The first snapshot is taken just before the first analysis

on January 5, 1993. The subsequent snapshots correspond to the 30th, 60th and 90th days of the

experiment (January 30, March 1 and March 31, 1993).

Note how the level of uncertainty associated with the depth of the thermocline layer changes in

the first month (between Figs. 1a and 1b). Thereafter, σ_{x} changes little for the equatorial

temperature except in the warm–pool area of the Western Pacific. In that area, the estimated

error goes down during the first month but increases between the 30th and 90th days although it

remains below its initial level. After 60 days, the estimated temperature–field uncertainty

approaches a steady state. Averaged over the entire equatorial section, the estimated error after

60 days is about 23% of the corresponding initial error estimate. The effect of the assimilation on the actual (*vs.* estimated) equatorial–temperature–field errors is examined in Section 3.4.

[Figure 1 near here, please]

3.3. Kalman gain vectors

Temperature analysis increments due solely to a $+1\,^{\circ}C$ sea–surface temperature (SST) innovation at $(180\,^{\circ}W,\ 0\,^{\circ}N)$ are shown in the vertical sections of Figure 2. This contribution is called marginal gain in this discussion because it measures the impact of processing a single measurement without reference to other data that are being assimilated at the same time.

The upper two panels show zonal and meridional sections through the marginal gain for the UOI. There, the marginal gain remains the same regardless how much data have been processed in previous analyses. Moreover, its structure is isotropic since the UOI operates under the assumption of a quasi–Gaussian $\mathbf{P}^f$.

The bottom six panels correspond to the MvEnKF. In this case, the marginal gain attributable to a unit innovation changes during the experiment. First, its amplitude is reduced to less than 20% of the initial amplitude after 30 days, as the filter weighs the model more and the observations less, after the model has presumably become more accurate following the assimilation of past measurements in earlier analyses. This is consistent with the error estimates shown in Figure 1, where σ_x is initially about $1\,^{\circ}C$ at the measurement location but drops to about $0.2\,^{\circ}C$ after 30 days. After 90 days, the marginal gain corresponding to this innovation has about the same

amplitude as after one month. The correlation structure also resembles the corresponding structure one month into the experiment, but there are subtle changes. These results and those of the previous Section suggest that, after some time, P^f may reach an asymptotic steady state when the observation distribution is quasi−steady. However, at this point, the MvEnKF has not been run for sufficiently long periods to verify whether such is the case.

[Figure 2 near here, please]

The marginal gain for salinity corresponding to the same unit SST innovation is shown in Figure 3. The six panels correspond to the bottom six panels of Figure 2. They exemplify how the MvEnKF provides a multivariate model for P^f to update model variables that are different from those being assimilated. Since the UOI only updates temperature, no panels in Figure 3 correspond to the upper two panels of Figure 2.

The dipole structure seen in the marginal temperature gain (surface: +, subsurface: −) is inverted in the marginal salinity gain. This inversion is consistent with intuition as, everything else being constant, increasing the water temperature and increasing its salt content have density− compensating effects on buoyancy. As in the case of the marginal temperature gain, the final amplitude of the marginal salinity gain is much less than the initial amplitude, reflecting the fact that the corresponding elements of σ_x have decreased accordingly.

[Figure 3 near here, please]

3.4. Equatorial temperature field

The monthly mean PMEL objective analysis of the equatorial–temperature field, *Teq,* is shown in Figure 4 for March 1993. The field shown there is a meridional average over the TAO moorings between $2°N$ and $2°S$ and is bilinearly interpolated in the (z, λ) plane (details available on PMEL website at http://www.pmel.noaa.gov/tao/). Figure 5 shows simulated monthly mean *Teq* fields for the same month averaged over the model grid cells at $1/3°N$ and $1/3°S$. The three panels correspond to the control run (top), the UOI run (middle) and the MvEnKF central forecast, x_c (bottom). The same contours and color scale are used in Figures 4 and 5.

In the control, the thermocline layer is too deep to the west and too shallow to the east, as can be seen by examining *T20*, the depth of the $20°C$ isotherm. The control run is also too warm near the surface in the warm–pool area, as is revealed by examination of *T28,* the $28°C$ isotherm depth. Overall, the UOI run is closer to the PMEL objective analysis than the control in terms of *T20.* However, in this run the onset of an instability is apparent in the high–zonal–wavenumber oscillations seen in the isotherms. This feature becomes noticeable after about a month of using the UOI to assimilate the TAO data. It is presumably a consequence of updating *T* while leaving *S, u* and *v* unchanged (see Troccoli *et al.* 2001), since it is absent from the control and from the MvEnKF run. In general, the MvEnKF does a better job in keeping *T, S u* and v in balance. This, in turn, results in a T_{eq} that also resembles the objective analysis of Figure 4 more than the corresponding fields from the control and UOI runs do. For example, *T20,* which is generally too deep in the control, especially to the west, is approximately at the same depth as in the PMEL analysis in the MvEnKF run. Also, to the east, *T20* is closer to the objective analysis in the

MvEnKF run than in the UOI run. The MvEnKF run also has a more accurate *T28* than either the UOI run or the control run does.

[Figures 4 and 5 near here, please]

The impact of the temperature assimilation on the model's hindcast skill is examined by continuing the integrations, now without assimilation, and validating against TAO–mooring observations between $2^{\circ}N$ and $2^{\circ}S$.   Time series of the mean, $\bar{y}$, and standard deviation, σ_y, of the difference, y, between the TAO–temperature observations and the corresponding model hindcasts are shown in Figure 6. To form the time series, the innovations, $d - l(x^f)$ in (1b), are calculated in the same manner as during the period with assimilation, but the corresponding data are not assimilated. The TAO data are considered error–free to simplify the discussion.

[Figure 6 near here, please]

Clearly, an important benefit of the temperature assimilation is that it corrects the forecast–model bias, resulting in $\bar{y}$ being initially nearer to zero in the MvEnKF and UOI runs than in the control (see KR01a). Thereafter, the control's *Teq* remains generally too warm, with mean differences larger that $0.5^{\circ}C$.   In contrast, the MvEnKF and UOI hindcast runs start with the spatial–mean residual, *i.e.* the hindcast's $\bar{y}$, close to zero. However, as no more data are assimilated, the mean residual becomes gradually more negative, indicating that the model drifts towards the overly warm conditions of the control. Yet, the reduction of the forecast bias following the TAO–temperature assimilation has a long–lasting effect. After 12 months of

forcing the model with climatological data and with no further assimilation, $\bar{y}$ is still closer to zero in the MvEnKF and UOI runs than in the control. Also, in this experiment, $\bar{y}$ is closer to zero during the hindcast when the MvEnKF is used for the initial ocean analysis, rather than the UOI. Thus, for March 31, 1993, x_c is a better initial condition than x_{oi} to predict *Teq* up to a year later.

Two important features of the MvEnKF are (1) the ability to update the entire model state, not just the observed quantity, and (2) the anisotropic, inhomogeneous, adaptive error estimation. To try to discern the relative importance of these two features, a univariate EnKF experiment was conducted to correct only *T* and not *S, u* and *v.* A comparison of *Teq* hindcasts for this experiment and the full MvEnKF shows that both are important. The *T*–only EnKF outperformed the UOI and the MvEnKF was marginally better than the univariate EnKF.

While $\bar{y}$ tracks the error component associated with the forecast–model bias, the σ_y time series shown in the bottom panel of Figure 6 can be interpreted as a diagnostic index for the unbiased error component. The net effect of the assimilation is to reduce this error component from about $2\,^{\circ}C$ in the control to about $1\,^{\circ}C$, when either the MvEnKF or the UOI is used to provide the initial ocean analysis for the hindcast.

Once the hindcasts have been initialized and no more measurements are assimilated, σ_y starts growing towards the control's level of unbiased error in both the MvEnKF and UOI runs. Moreover, the two assimilation methods are indistinguishable from the point of view of the σ_y statistic.

In summary, the impact of processing the data with the MvEnKF rather than with the UOI is more noticeable on $\bar{y}$, the biased error component, than on σ_y, the unbiased .component Because it simultaneously updates *T, S, u* and *v,* the MvEnKF maintains a better balance between the model prognostic fields. The better balance results in a longer lived forecast–bias correction, thereby increasing the skill of the model *Teq* hindcasts on SI timescales. Both methods have similar impacts on the unbiased error component.

3.5. *Equatorial zonal current*s

Notwithstanding some interruptions, current–meter data corresponding to the period spanned by the experiment are available from four ADCP–equipped equatorial TAO moorings at $165°E$, $170°W$, $140°W$ and $110°W$. The profilers sample *u* and *v* vertically from the surface down to about $1000 m$. The top two panels in Figure 7 show time–depth diagrams of *u* at the $170°W$ and $140°W$ moorings from January 1 to March 31, 1993, *i.e.* during the three months with temperature assimilation. The second–row panels show the corresponding diagrams for the control run. The third and fourth rows correspond to the UOI and MvEnKF runs, respectively.

[Figure 7 near here, please]

At $170°W$, the undercurrent flows at a depth of about $200 m$. There, the maximum current speed reached during January–March, 1993 is $0.81 ms^{-1}$, $0.93 ms^{-1}$ and $0.97 ms^{-1}$ in the control, UOI and MvEnKF runs, respectively. In the observations, *u* reaches a maximum of $0.98 ms^{-1}$.

One problem seen in the UOI run, but not in the control or in the MvEnKF run, is that the undercurrent at $170^{\circ}W$ flows westward from March 15, 1993 on. As in the previous Section, improper balancing of the model's T and u fields resulting from only the former being updated is the most likely cause of this unrealistic behavior.

At $140^{\circ}W$, the observed undercurrent strength at a depth of about $120\,m$ weakens during February 1993 and then increases in the next month. The control completely misses this change. It is present in both the MvEnKF and UOI runs, but the MvEnKF run resembles the observations more closely. For example, the minimum current speed at this depth during the first half of February 1993 is $0.49\,ms^{-1}$, $0.19\,ms^{-1}$ and $0.37\,ms^{-1}$ in the control, UOI and MvEnKF, in that order, while it is $0.40\,ms^{-1}$ in the ADCP data.

To gain better insight into how the assimilation impacts the model currents, Figure 8 shows the time series of the correlation, c_u, between the simulated vertical u profiles and the corresponding ADCP−u observations, at each ADCP−equipped mooring and during the period with temperature assimilation. Overall, c_u is the highest in the MvEnKF run, except at $165^{\circ}E$, where it is initially higher in the UOI run, but the MvEnKF, UOI and control are all poorly correlated with the observations during the third month. At $110^{\circ}W$, c_u in the MvEnKF run is already close to one after the first temperature analysis. The MvEnKF run and the data are also highly correlated during the entire 90−day assimilation period at $170^{\circ}W$.

[Figure 8 near here, please]

18

Figure 9 shows time–depth diagrams similar to those of figure 7, but this time for April 1, 1993– June 30, 1993, that is, for the first 90 days of the no–assimilation hindcast runs. The zonal current at $170°W$ and at $140°W$ from the UOI run during this period least resembles the observations. This is especially true during April 1993, after cessation of the assimilation and initialization of the hindcasts. Evidently, it takes a while for the transients associated with the unrealistic initial states to die out and for the forcing to reestablish the more realistic temperature and flow structures along the Equator.

[Figure 9 near here, please]

At $170°W$, the undercurrent in the UOI run flows westward in the first month before reversing direction at the beginning of May 1993. From then on, the UOI's undercurrent at this location matches the observations reasonably well, although the local acceleration observed in mid May– mid June appears early in this run.

At $140°W$, the UOI underestimates the undercurrent strength during the hindcast run's first month. This is less problematic in the control and MvEnKF runs. Yet the control's undercurrent is overly diffuse. The MvEnKF run is better than the control in this respect.

While the temperature assimilation has a positive impact on the T–field hindcasts at lead times of up to one year, its impact on the u–field hindcasts in this experiment becomes statistically hard to discern from the control after three months. This observation is pertinent to both the UOI and the MvEnKF runs and may reflect several factors.

19

First, after about three months of not assimilating anything, any imbalances between the initial T and u fields have been compensated by the model. From then on, the u field from the UOI run is about as good as that from the control. Second, the *Teq* hindcast–skill increase following the TAO temperature assimilation with either the UOI or the MvEnKF is mostly due to a lessening of the forecast bias. As seen in the preceding Section, the model remembers the bias correction longer than it remembers the correction to the unbiased T–error component, σ_y. Since the model currents are generally unbiased, the impact of the temperature assimilation on u disappears at about the same rate as its impact on σ_y.

4 Conclusions

This paper's main objective is to ascertain whether the MvEnKF can realistically be used in place of standard OI methodologies for the purpose of ocean initialization and as part of a coupled SI forecasting system. With these questions in mind, *in situ* TAO temperature data are processed and the impact of their assimilation on the simulated temperature and currents is examined.

Naturally, more experiments are needed before the tentative conclusions drawn herein can be generalized. Still, the discussion of Section 3 has helped identify a number of issues. Of these, the need to account for the fact that the model errors are not unbiased, as the theory of the Kalman filter assumes, is of crucial importance because it directly impacts the skill of the coupled–model hindcasts (Vintzileos *et al.* 2001). In response, an algorithm based on Dee and DaSilva (1998), to estimate the unbiased and biased model–error components separately, has recently been included in the NSIPP data assimilation system and is currently being tested.

The presence of a forecast bias introduces additional difficulties when there are areas without observations. For example, the continuous assimilation of the TAO temperature data without properly specifying the model–error statistics is bound to introduce discontinuities or sudden changes in the forecasts away from the Equator, since the assimilation corrects the temperature–field bias in the inner Tropics but not elsewhere. Therefore, it is important to consider the joint assimilation of various data types and to use observations with global coverage if available, even when focusing on a limited region such as the Tropical Pacific. The MvEnKF is ideally suited for such applications because it dynamically estimates multivariate error statistics and considers their inhomogeneity and anisotropy in the calculation of the Kalman gain matrix.

There are many outstanding issues. Among these, the main technical ones are related to computational cost, ensemble initialization and process–noise modeling, and are discussed in KR01a and KR01b. Another important issue is whether to perform the assimilation on constant–depth levels, as is done here, or within the framework of the isopycnal model formulation, *i.e.* without switching back and forth between the "layer" and "level" state–vector representations. Recent experiments with the MvEnKF and with the related MvOI system also developed at NSIPP have established that this choice can have significant repercussions on the outcome of the assimilation.

References

Atlas, R., R. Hoffman, S. Bloom, J. Jusem, and J. Ardizzone, 1996: A multiyear global surface wind velocity dataset using SSM/I wind observations. *Bull. Amer. Met. Soc.*, **77**, 869−882.

Borovikov A., and M. Rienecker, 2001: Multivariate error covariance estimates by Monte−Carlo simulation for Pacific Ocean assimilation studies. draft manuscript.

Borovikov, A., M. Rienecker, and P. Schopf, 2001: Surface heat balance in the Equatorial Pacific Ocean: climatology and the warming event of 1994−95. *J. Clim.*, **14**, 2624−2641.

Burgers, G., P. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719−1724.

Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457pp.

Dee, D., and A. Da Silva, 1998: Data assimilation in the presence of forecast bias. *Q. J. R. Meteorol. Soc.*, **124**, 269−295.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi−geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, C**99**, 10,143−10,162.

Gaspari, G., and S. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723−757.

Gelb, A. (Ed.), 1974: *Applied Optimal Estimation*. MIT Press, 374pp.

Ghil, M., and P. Malanotte−Rizzoli, 1991: Data assimilation in meteorology and oceanography. *Adv. Geophys.*, **33**, 141−266.

Houtekamer, P., and H. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123−137.

Ji, M., and A. Leetmaa, 1997: Impact of data assimilation on ocean initialization and El Niño prediction. *Mon. Wea. Rev.*, **125**, 742−753.

Keppenne, C., 2000: Data assimilation into a primitive−equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 1971−1981.

Keppenne, C., and M. Rienecker, 2001a: Development and initial testing of a parallel ensemble Kalman filter for the Poseidon isopycnal ocean general circulation model. *Mon. Wea. Rev.*, submitted.

Keppenne, C., and M. Rienecker, 2001b: *Design and Implementation of a Parallel Multivariate Ensemble Kalman Filter for the Poseidon Ocean General Circulation Model.* NASA Technical Report Series on Global Modeling and Data Assimilation, Vol. 19, M. Suarez, ed., 31pp.

Konchady, M., A. Sood, and P. Schopf, 1998: Implementation and performance evaluation of a parallel ocean model. *Parallel Comput.*, **24**, 181−203.

McPhaden, M., A. Busalacchi, R. Cheney, J. Donguy, K. Gage, D. Halpern, M. Ji, P. Julian, G. Meyers, G. Mitchum, P. Niiler, J. Picaut, R. Reynolds, N. Smith, and K. Takeuchi, 1998: The Tropical Ocean−Global Atmosphere observing system: a decade of progress. *J. Geophys. Res.,* **C103**, 14169−14240.

Mitchell, H., and P. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 416−433.

Reynolds, R., and T. Smith, 1994: Improved global sea−surface temperature analyses using optimum interpolation. *J. Climate*, **7**, 929−948.

Schaffer, D., and M. Suarez, 1998: Next stop: teraflop; the parallelization of an atmospheric general circulation model. Available from http://nsipp.gsfc.nasa.gov/ .

Schopf, P., and A. Loughe, 1995: A reduced–gravity isopycnic ocean model–hindcasts of El–Niño. *Mon. Wea. Rev.*, **123**, 2839–2863.

Troccoli, A., M. Rienecker, C. Keppenne, and G. Johnson, 2001: Temperature data assimilation with salinity corrections: validation in the tropical Pacific Ocean, 1996–1998. *J. Geophys. Res.*, submitted.

Vintzileos, A., M. Rienecker, M. Suarez, S. Milleri, and A. Borovikov, 2001: Prediction of Tropical Pacific SST for 2001 with the NSIPP ENSO forecasting system. Paper presented at the 2001 Climate Diagnostics Workshop held at La Jolla, CA, October 22–25, 2001.

Yang, S., K. Lau and P. Schopf, 1999: Sensitivity of the tropical Pacific Ocean to precipitation induced freshwater flux, *Clim. Dynam.*, 15, 737–750.

Yuan, D., M. Rienecker, and P. Schopf, 2001: Nonlinear reflection of the equatorial Rossby waves at the Pacific western boundary and its role in ENSO. *J. Clim.*, submitted.
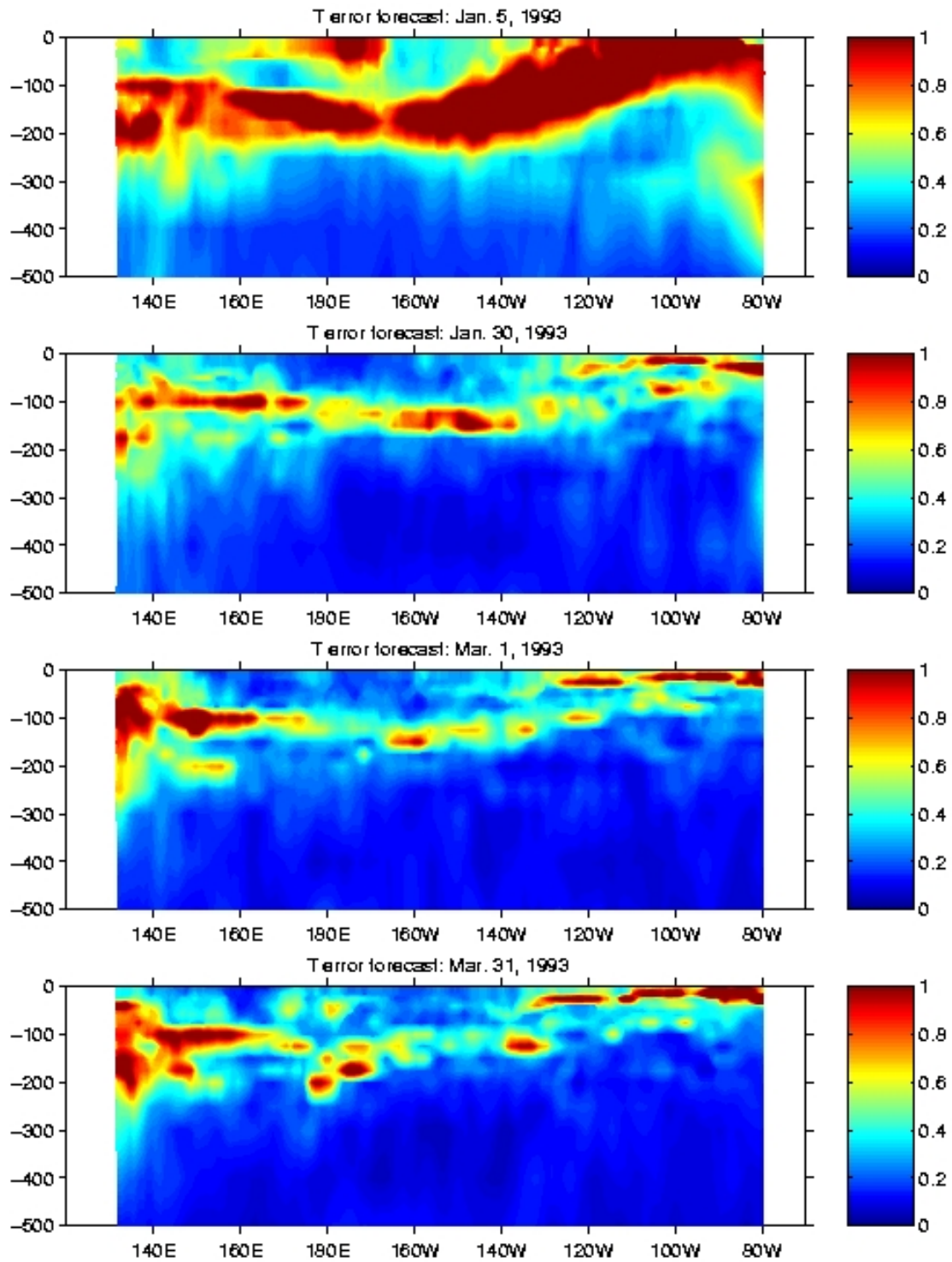
Figure 1. Estimated temperature–field background–error along the Equator. From top to bottom: prior to the first analysis, after 30 days, after 60 days and after 90 days.
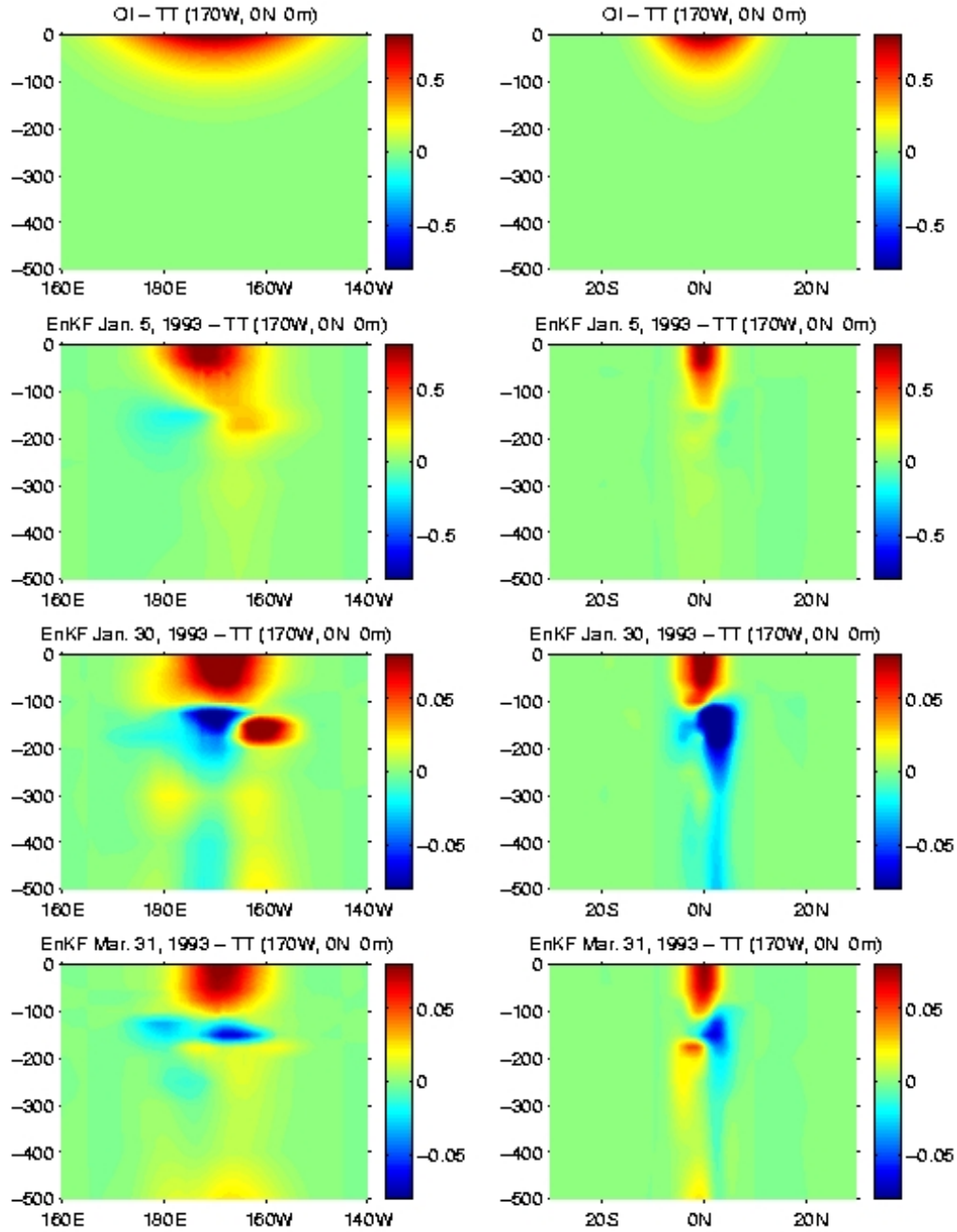
Figure 2. Contribution of a $+1°C$ innovation at $(180°W, 0°N)$ to the temperature−field part of the analysis increments (marginal temperature gain in the text). Zonal sections (left) and meridional sections (right) are shown. From top to bottom: UOI, MvEnKF prior to first analysis, MvEnKF after 30 days and MvEnKF after 90 days.
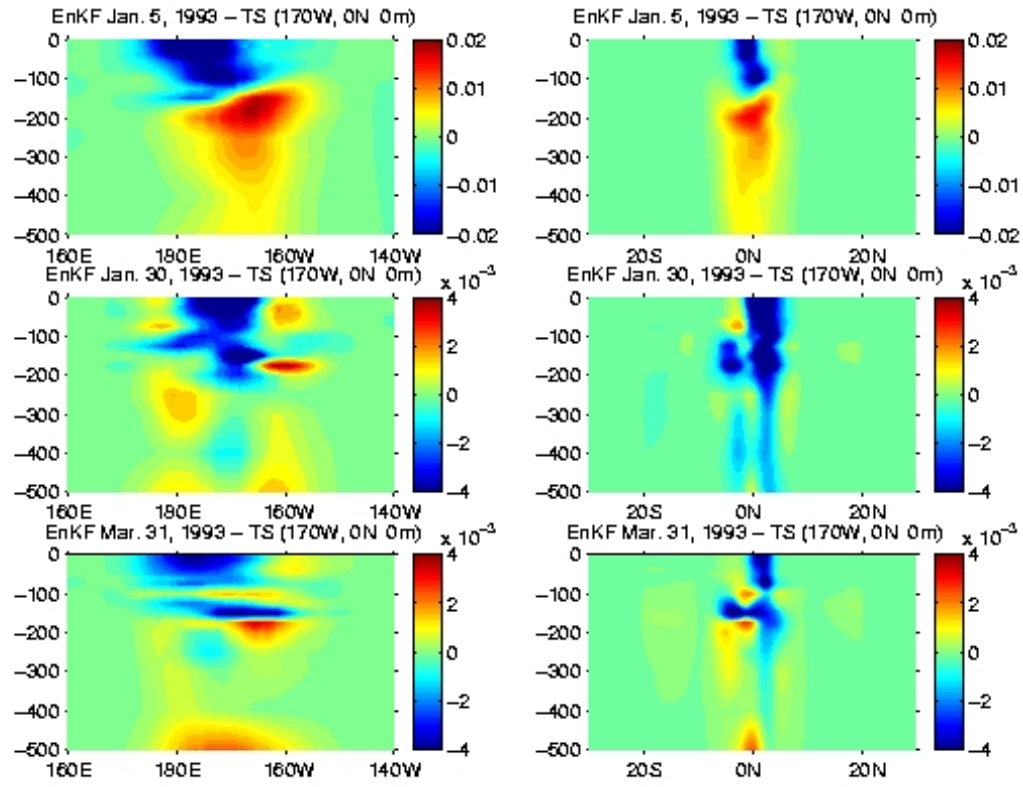
Figure 3. Same as the bottom six panels of Figure 2 for the marginal salt gain.
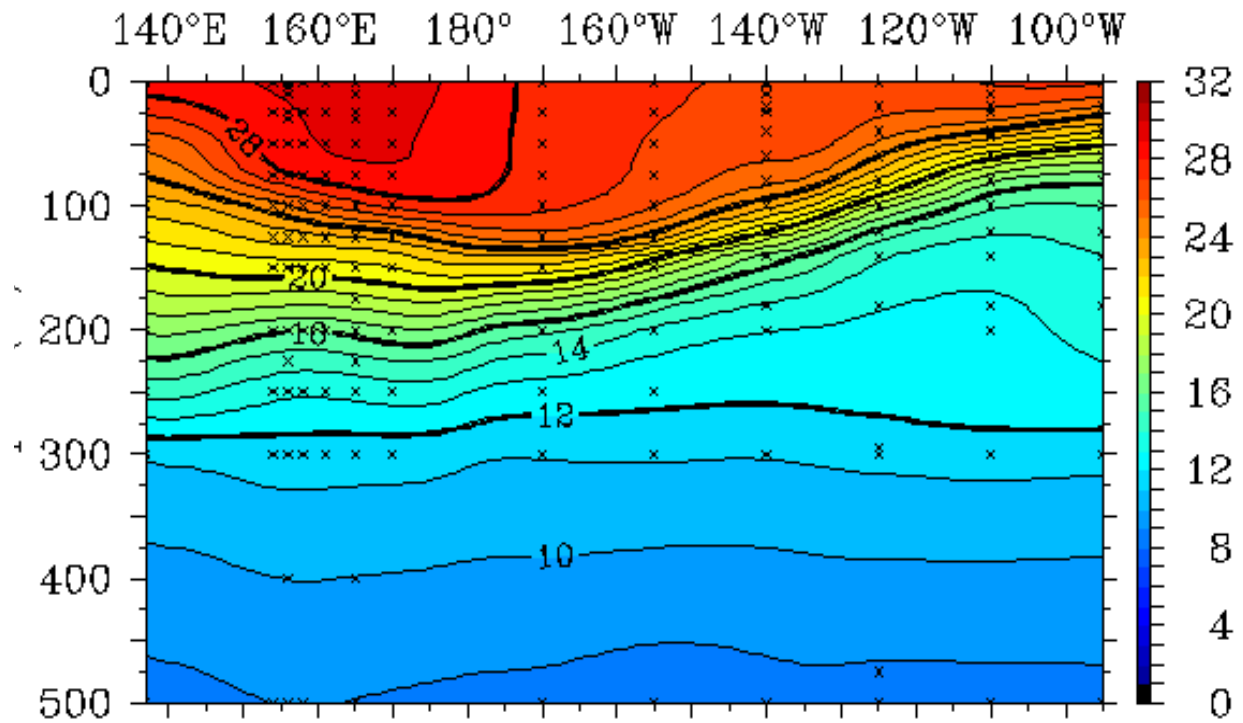
Figure 4. NOAA/PMEL monthly mean objective analysis of TAO temperature data between $2°N$ and $2°S$ for March 1993. From PMEL website (http://www.pmel.noaa.gov/tao/).
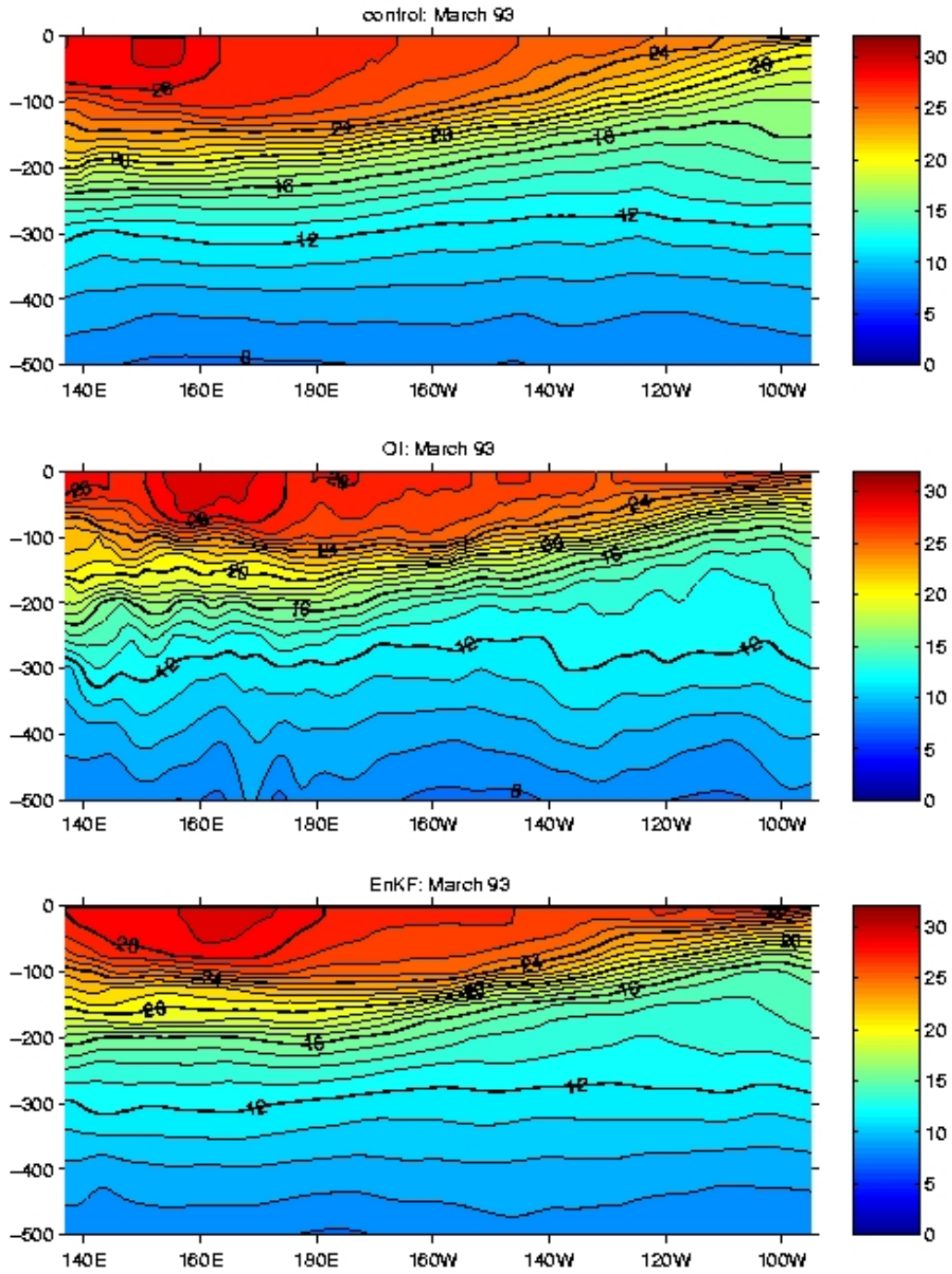
Figure 5. Same as Figure 4 for simulated temperature field in TAO–temperature assimilation experiment. From top to bottom: control, UOI and MvEnKF. Refer to the text for detailed explanation.
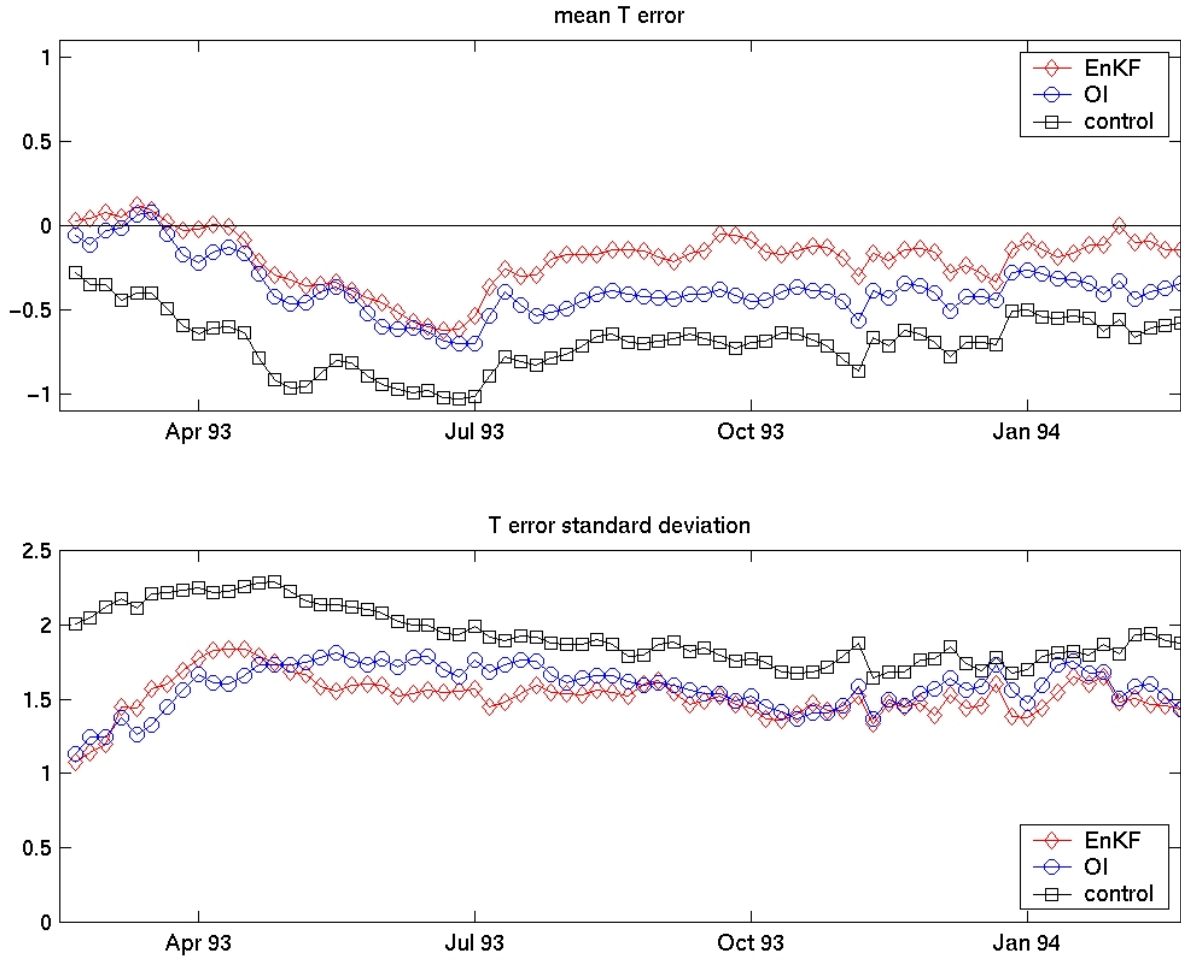
Figure 6. Time series of (top) Mean, $\bar{y}$, and (bottom) standard deviation, σ_y, of observations minus forecast residual between $2^\circ N$ and $2^\circ S$ during climatologically forced hindcasts without assimilation. The diamonds, circles and squares identify the MvEnKF, UOI and control, respectively.
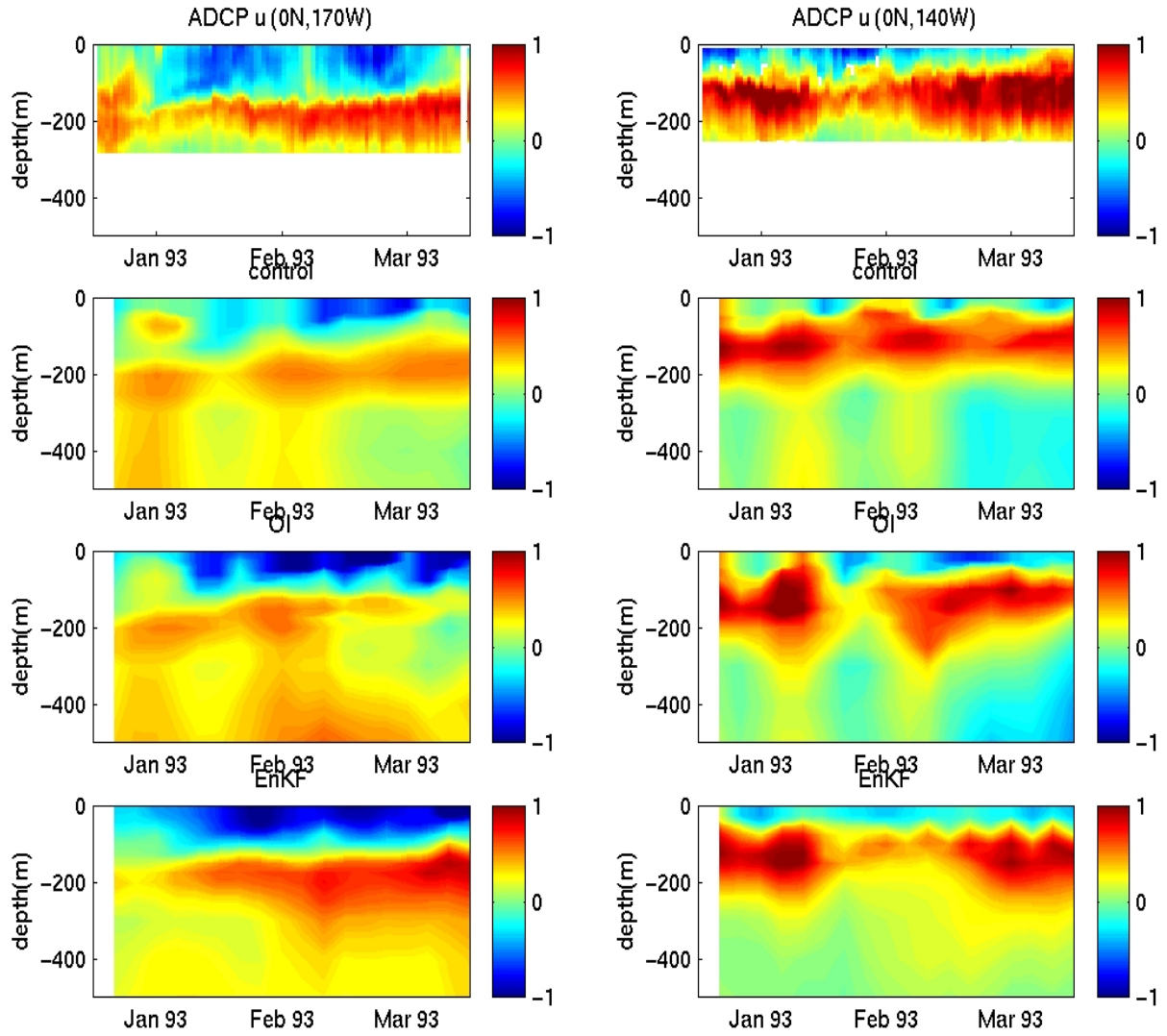
Figure 7. Time–depth diagrams of the zonal–current component at (left) $170°W$ and (right) $140°W$ and along the Equator during experiment phase with TAO–temperature assimilation. From top to bottom: ADCP data, control, UOI and MvEnKF. The horizontal thickmarks correspond to the 15th of the month indicated.
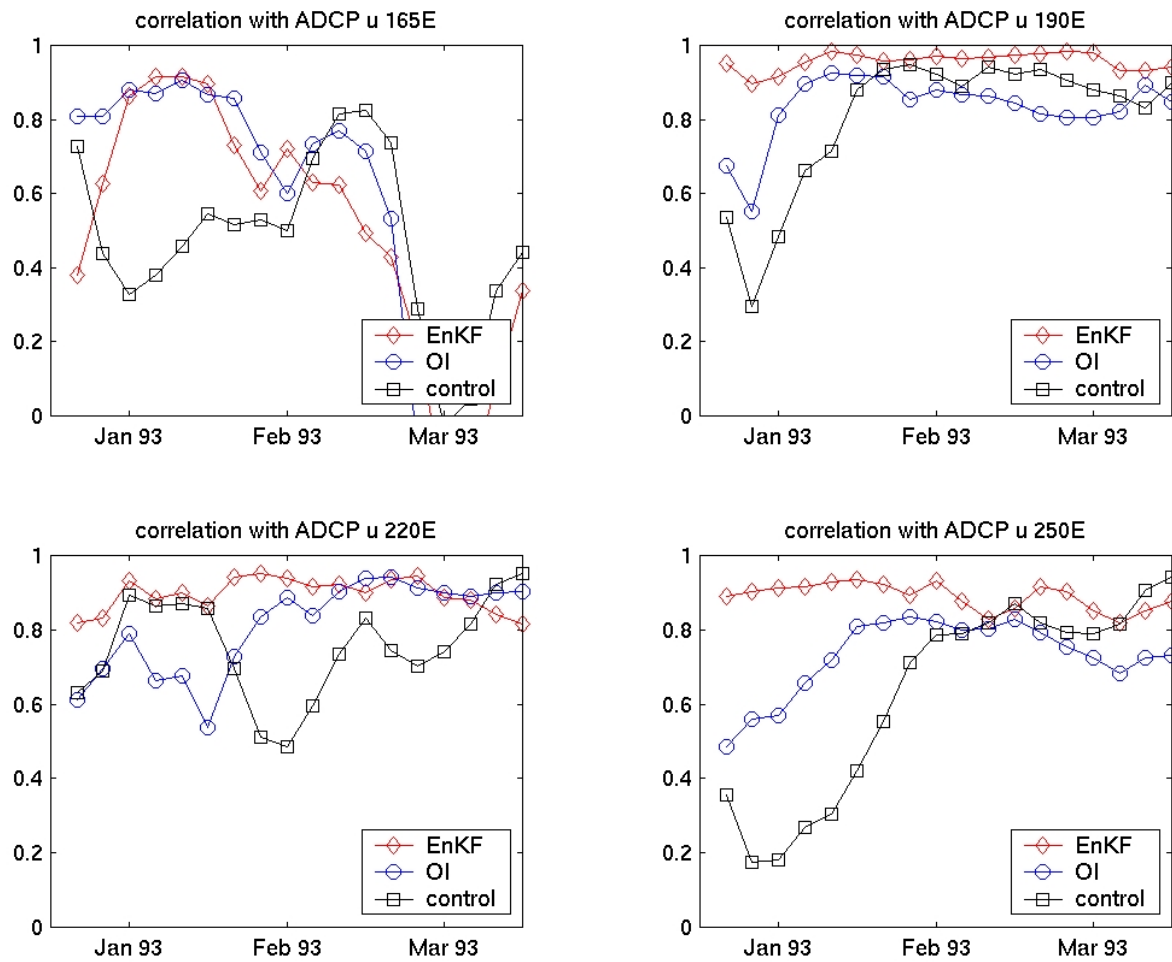
Figure 8. Time evolution of correlation with vertical profiles of ADCP–zonal–current observations at (upper left) $165^{\circ}E$, (upper right) $170^{\circ}W$, (lower left) $140^{\circ}W$ and (lower right) $110^{\circ}W$ in the MvEnKF (diamonds), the UOI (circles) and the control (squares).
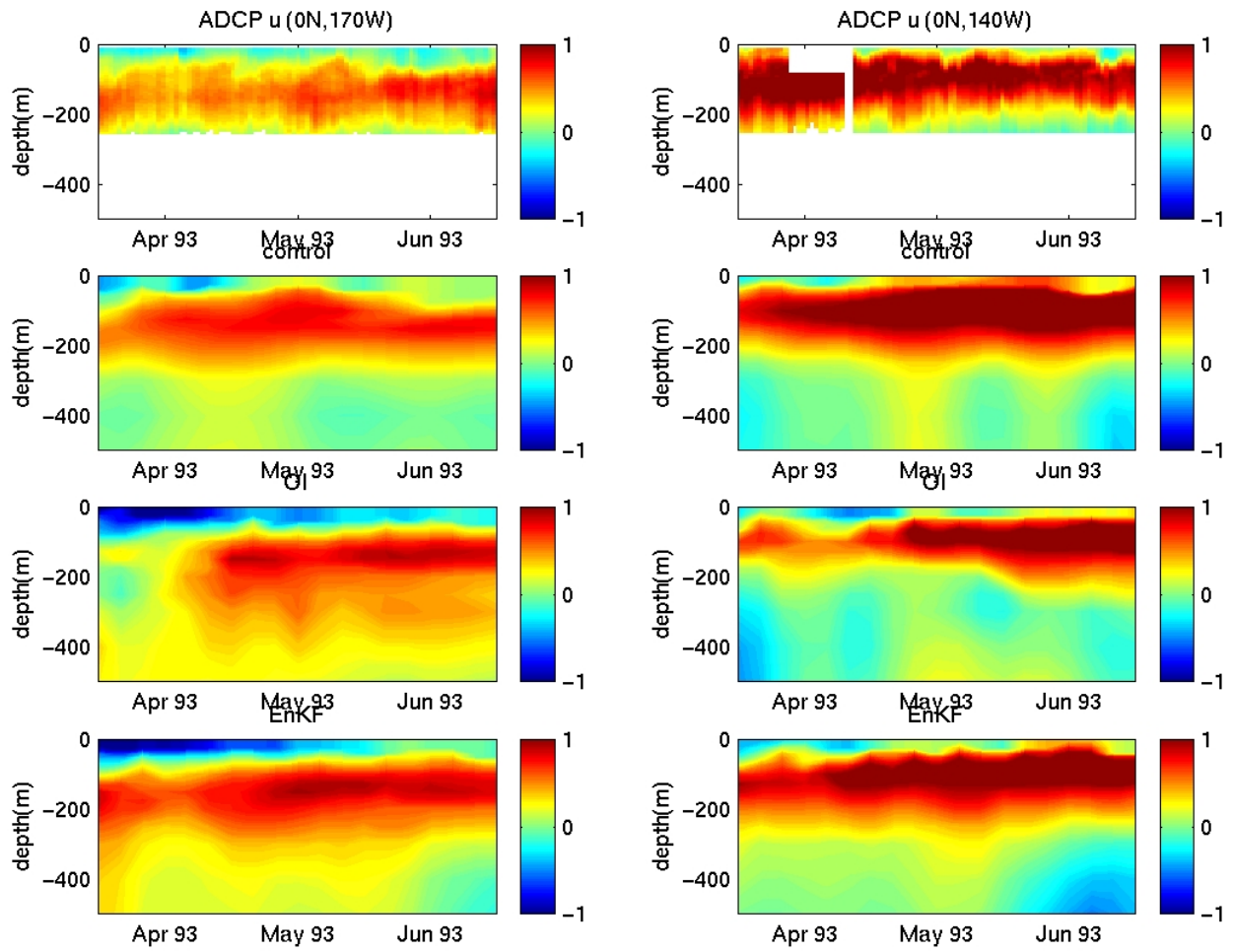
Figure 9. Same as Figure 7 for the first 90 days of the hindcast phase of the experiment in which no data assimilation takes place.